

# Concentration of information content for convex measures

Matthieu Fradelizi\*, Jiange Li<sup>†</sup> and Mokshay Madiman<sup>‡</sup>

## Abstract

We obtain sharp exponential deviation estimates of the information content as well as a sharp bound on the varentropy for convex measures, extending the development for log-concave measures in recent work of Fradelizi, Madiman and Wang (2015).

**Note:** This document is not intended for submission to a journal in its current form. It will form one part of a larger paper [1] (which is in preparation), and is being made available separately for quicker accessibility for those interested.

## 1 Introduction

Let  $X$  be a random variable taking values in  $\mathbb{R}^n$ . Suppose that the distribution of  $X$  has density  $f$  with respect to the Lebesgue measure on  $\mathbb{R}^n$ . The information content of  $X$  is the random variable

$$\tilde{h}(X) = -\log f(X). \quad (1)$$

The average value of  $\tilde{h}(X)$  is known more commonly as the entropy. The entropy of  $X$  is defined by

$$h(X) = -\int_{\mathbb{R}^n} f(x) \log f(x) dx = -\mathbb{E} \log f(X). \quad (2)$$

Because of the relevance of the information content in various areas such as information theory, probability and statistics, it is intrinsically interesting to understand its behavior. In particular, it is natural to ask whether the information content concentrates around the entropy in high dimension. If  $X$  is a standard Gaussian random vector in  $\mathbb{R}^n$ , its information content is

$$\tilde{h}(X) = \frac{|X|^2}{2} + \frac{n}{2} \log(2\pi),$$

where  $|\cdot|$  is the Euclidean norm. In this case, the concentration property of  $|X|^2$  around its mean is well known. Bobkov and Madiman [2] showed that  $\tilde{h}(X)$  does possess a powerful concentration property if  $X$  has a log-concave density. That in particular includes the Gaussian

---

\*Laboratoire d'Analyse et de Mathématiques Appliquées UMR 8050, Université Paris-Est Marne-la-Vallée, 5 Bd Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France. E-mail: matthieu.fradelizi@u-pem.fr

<sup>†</sup>Department of Mathematical Sciences, University of Delaware, 501 Ewing Hall, Newark, DE 19716, USA. E-mail: lijange@udel.edu

<sup>‡</sup>Department of Mathematical Sciences, University of Delaware, 501 Ewing Hall, Newark, DE 19716, USA. E-mail: madiman@udel.edu

case, which had earlier been treated by Cover and Pombra [9]. The proof heavily depends on the localization lemma of Lovász-Simonovits [15] and reverse Hölder type inequalities [6]. As a consequence, the following variance bound holds for all log-concave random variables

$$\text{Var}(\tilde{h}(X)) \leq Cn \quad (3)$$

with some absolute constant  $C$ . Recently, it was independently determined by Nguyen [17] and Wang [18] that the sharp constant is 1. Simpler proofs of the sharp constant are independently given by Bolley, Gentil and Guillin [5] and Fradelizi, Madiman and Wang [11]. The key observation in [11] is that variance bound emerges as a consequence of the log-concavity of the moments of log-concave functions.

In this note, we extend the concentration property of the information content from log-concave measures to general convex measures. In Section 2, we will show that exponential deviation of a functional follow from the log-concavity of normalized moments of that functional. In particular, the log-concavity of the normalized moments of  $s$ -concave functions are studied in Section 3. Optimal concentration and sharp variance bound of the information content of  $\kappa$ -concave random variables are obtained in Section 4.

## 2 A general principle for exponential deviation

Let  $X$  be a random variable taking values in  $\mathbb{R}^n$ . Suppose that it has density  $f$  with respect to the Lebesgue measure on  $\mathbb{R}^n$ . Let  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  be real-valued function. One natural way to show the exponential deviation of  $\varphi(X)$  from its mean is to prove the finiteness of the moment generating function  $\mathbb{E}e^{\alpha\varphi(X)}$  for certain  $\alpha$ . The logarithmic moment generating function  $L(\alpha)$  is defined by

$$L(\alpha) = \log \mathbb{E}e^{\alpha\varphi(X)}. \quad (4)$$

The following observation is a well known fact about exponential families in statistics. Let  $a, b > 0$  be certain real numbers.

**Lemma 2.1.** *Suppose that  $L(\alpha) < \infty$  for  $\alpha \in (-a, b)$ . Then we have*

$$L'(\alpha) = \mathbb{E}\varphi(X_\alpha), \quad L''(\alpha) = \text{Var}(\varphi(X_\alpha)), \quad (5)$$

where  $X_\alpha$  is a random variable with density

$$f_\alpha(x) = \frac{e^{\alpha\varphi(x)} f(x)}{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) dx}. \quad (6)$$

For  $\alpha = 0$ , one particularly has  $L'(0) = \mathbb{E}\varphi(X)$  and  $L''(0) = \text{Var}(\varphi(X))$ .

*Proof.* The assumption  $L(\alpha) < \infty$  for  $\alpha \in (-a, b)$  guarantees that  $L(\alpha)$  is infinitely differentiable with respect to  $\alpha \in (-a, b)$  and that we can freely change the order of differentiation and expectation. Then we have

$$L'(\alpha) = \frac{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) \varphi(x) dx}{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) dx} = \mathbb{E}\varphi(X_\alpha).$$

Differentiate  $L'(\alpha)$  one more time. We have

$$L''(\alpha) = \frac{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) \varphi^2(x) dx}{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) dx} - \left( \frac{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) \varphi(x) dx}{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)} f(x) dx} \right)^2 = \text{Var}(\varphi(X_\alpha)).$$

□

**Definition 2.2.** A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is called *log-concave* if we have

$$f((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda} f(y)^\lambda \quad (7)$$

for all  $x, y \in \mathbb{R}^n$  and all  $\lambda \in [0, 1]$ .

The following lemma tells us that the upper bound of  $\mathbb{E}e^{\alpha\varphi(X)}$  emerges as a consequence of the log-concavity of  $L(\alpha)$  after certain normalization.

**Lemma 2.3.** Let  $c(\alpha)$  be a smooth function such that  $e^{-c(\alpha)}\mathbb{E}e^{\alpha\varphi(X)}$  is log-concave for  $-a < \alpha < b$ . Then we have

$$\mathbb{E}e^{\alpha(\varphi(X) - \mathbb{E}\varphi(X))} \leq e^{\psi_c(\alpha)}, \quad (8)$$

where  $\psi_c(\alpha) = c(\alpha) - c(0) - c'(0)\alpha$ .

*Proof.* Since  $e^{-c(s)}\mathbb{E}e^{s\varphi(X)}$  is log-concave, we have  $L''(s) \leq c''(s)$ . For any  $0 < t < \alpha < b$ , integrating the inequality over  $(0, t)$  we have

$$L'(t) - L'(0) \leq c'(t) - c'(0).$$

Integrating both sides over  $(0, \alpha)$ , we have

$$L(\alpha) - L(0) - L'(0)\alpha \leq c(\alpha) - c(0) - c'(0)\alpha. \quad (9)$$

Similarly we can show that the estimate also holds for  $-a < \alpha < 0$ . Notice that  $L(0) = 0$  and  $L'(0) = \mathbb{E}\varphi(X)$ . Then the lemma follows from exponentiating both sides of (9).  $\square$

**Remark 2.4.** From Lemma 2.1 and Lemma 2.3, we can see that the study of upper bound of  $\text{Var}(\varphi(X_\alpha))$  is equivalent to that of the normalizing function for  $\mathbb{E}e^{\alpha\varphi(X)}$  to be log-concave. We can get one from the other by differentiating or integrating twice. That is why variance bounds can imply exponential deviation inequalities when moment generating functions exist.

Let  $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$  be a real-valued function. For  $x \in \mathbb{R}$ , the Fenchel-Legendre dual function  $f^*(x)$  is defined by

$$f^*(x) = \sup_y (xy - f(y)). \quad (10)$$

Let  $\psi_{c,+}(\alpha)$  and  $\psi_{c,-}(\alpha)$  be the restrictions of  $\psi_c(\alpha)$  on the positive and negative half axis, respectively.

**Corollary 2.5.** Under the assumptions and notations of Lemma 2.3, we have for any  $t > 0$  that

$$\mathbb{P}(\varphi(X) - \mathbb{E}\varphi(X) > t) \leq e^{-\psi_{c,+}^*(t)}, \quad (11)$$

$$\mathbb{P}(\varphi(X) - \mathbb{E}\varphi(X) < -t) \leq e^{-\psi_{c,-}^*(-t)}, \quad (12)$$

where  $\psi_{c,+}^*$ ,  $\psi_{c,-}^*$  are Fenchel-Legendre dual functions of  $\psi_{c,+}$ ,  $\psi_{c,-}$ , respectively.

*Proof.* The proof follows from the so-called Cramér-Chernoff method: using Markov inequality in conjugation with optimization of the resulting bound. For the upper tail, we have for  $0 < \alpha < b$  and  $t > 0$  that

$$\begin{aligned}\mathbb{P}(\varphi(X) - \mathbb{E}(\varphi(X)) > t) &= \mathbb{P}(e^{\alpha(\varphi(X) - \mathbb{E}(\varphi(X)))} > e^{\alpha t}) \\ &\leq e^{-\alpha t} \cdot \mathbb{E}e^{\alpha(\varphi(X) - \mathbb{E}(\varphi(X)))} \\ &\leq e^{-(\alpha t - \psi_{c,+}(\alpha))}.\end{aligned}$$

We use Lemma 2.3 in the second inequality. Then the upper tail estimate follows by taking the infimum of the right hand side over  $0 < \alpha < b$ . The lower tail estimate follows from the same argument for  $-a < \alpha < 0$ .  $\square$

### 3 Log-concavity of the moments of $s$ -concave functions

In this section, we study the log-concavity of the (normalized) moments of  $s$ -concave functions, which, in conjugation with the results from the previous section, will enable us to obtain optimal concentration of the information content for convex measures.

**Definition 3.1.** For  $s \in \mathbb{R}$ , a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is called  $s$ -concave if we have

$$f((1 - \lambda)x + \lambda y) \geq ((1 - \lambda)f(x)^s + \lambda f(y)^s)^{1/s} \quad (13)$$

for all  $x, y$  such that  $f(x)f(y) > 0$  and for all  $\lambda \in [0, 1]$ .

For  $s = 0$ , the right hand side is defined by continuity, which corresponds to log-concave functions defined before. For  $s > 0$ , the previous definition is equivalent to that  $f^s$  is concave on its support; while for  $s < 0$ , it is equivalent to that  $f^s$  is convex on its support.

Recall that for  $x > 0$ , the gamma function  $\Gamma(x)$  is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

For  $x, y > 0$ , the beta function  $B(x, y)$  is defined by

$$B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt.$$

The following result is proved by Borell [6] for  $s > 0$ , except that the function  $\varphi$  is assumed to be decreasing. It was then noticed by some people and available for example in Guédon, Nayar and Tkocz [12] that the result remains true without any monotonicity hypothesis. For  $s < 0$ , it is proved by Fradelizi, Guédon and Pajor [10], and the case  $s = 0$  follows by taking the limits (or reproducing the mechanics of the proof).

**Proposition 3.2.** Let  $s \in \mathbb{R}$  and let  $\varphi : [0, +\infty) \rightarrow [0, +\infty)$  be an  $s$ -concave integrable function.

- 1) If  $s > 0$ , then  $p \mapsto B(p, s^{-1} + 1)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$  is log-concave for  $p > 0$ .
- 2) If  $s = 0$ , then  $p \mapsto \Gamma(p)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$  is log-concave for  $p > 0$ .
- 3) If  $s < 0$ , then  $p \mapsto B(p, -s^{-1} - p)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$  is log-concave for  $0 < p < -1/s$ .

Let us define the function  $\varphi_s(t) = (1 - st)_+^{1/s} \mathbf{1}_{\mathbb{R}_+}$  for  $s \neq 0$ , and  $\varphi_0(t) = e^{-t} \mathbf{1}_{\mathbb{R}_+}$ . Then the preceding proposition may be expressed in the following way: if  $\varphi : [0, +\infty) \rightarrow [0, +\infty)$  is  $s$ -concave, then the function

$$p \mapsto \frac{\int_0^{+\infty} t^{p-1} \varphi(t) dt}{\int_0^{+\infty} t^{p-1} \varphi_s(t) dt}$$

is log-concave for  $p$  such that  $1/p > \max(0, -s)$ . Using the preceding proposition, we can prove the following theorem which unifies and partially extends previous results of Borell [6], Bobkov and Madiman [3], and Fradelizi, Madiman and Wang [11]. A weaker log-concavity statement was also obtained by Nguyen[16].

**Theorem 3.1.** *Let  $s \in \mathbb{R}$  and let  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  be an integrable  $s$ -concave function. Then the function*

$$p \mapsto (p + s) \cdots (p + ns) \int_{\mathbb{R}^n} f(x)^p dx \quad (14)$$

*is log-concave for  $p > \max(0, -ns)$ .*

*Proof.* The case  $s = 1$  is due to Borell [6] and the case  $s > 0$  deduces directly by applying Borell's result to  $f^s$ . The case  $s = 0$  was proved by Fradelizi, Madiman and Wang [11]. The case  $s = -1$  is due to Bobkov and Madiman [3]<sup>1</sup>, except that the range was  $p > n + 1$ . In the same way, the case  $s < 0$  deduces from the case  $s = -1$  by applying it to  $f^{|s|}$ . So we only need to prove the extension of the range for  $s = -1$ . Let us assume that  $s = -1$ . Thus  $f$  is  $-1$ -concave, which means that  $g = f^{-1}$  is convex on its support. As done by Bobkov and Madiman [3], we write

$$\int_{\mathbb{R}^n} f(x)^p dx = \int_{\mathbb{R}^n} g(x)^{-p} dx = \int_0^{+\infty} p t^{p-1} \psi(1/t) dt,$$

where  $\psi(t) = |\{x \in \mathbb{R}^n : g(x) \leq t\}|_n$  is the Lebesgue measure of the sub-level set  $\{x \in \mathbb{R}^n : g(x) \leq t\}$ . Using Brunn-Minkowski theorem, we can see that  $\psi$  is a  $1/n$ -concave function. Using the properties of the perspective function, we can deduce that the function  $\varphi(t) = t^n \psi(1/t)$  is also a  $1/n$ -concave function. Thus it follows that

$$\int_{\mathbb{R}^n} f(x)^p dx = p \int_0^{+\infty} t^{p-n-1} \varphi(t) dt.$$

Applying Proposition 3.2 to  $s = 1/n$  and  $p$  replaced by  $p - n$  we get that

$$B(p - n, n + 1)^{-1} \int_0^{+\infty} t^{p-1-n} \varphi(t) dt$$

is log-concave on  $(n, +\infty)$ . Then we can conclude the proof using the following identity

$$B(p - n, n + 1)^{-1} = \frac{p(p-1) \cdots (p-n)}{\Gamma(n+1)}.$$

□

---

<sup>1</sup>The details of this proof were omitted from [3] because of space considerations, and are being presented here. A complete presentation will appear in [1].

The fact that Theorem 3.1 is optimal can be seen from the following example. Let  $U : \mathbb{R}^n \rightarrow [0, \infty]$  be a positively homogeneous convex function of degree 1, i.e. that  $U(tx) = tU(x)$  for all  $x \in \mathbb{R}^n$  and all  $t > 0$ . We define  $f_{s,U} = (1 - sU)_+^{1/s}$  for  $s \neq 0$  and  $f_{0,U} = e^{-U}$  for  $s = 0$ . Then we have

$$\int_{\mathbb{R}^n} f_{s,U}(x)^p dx = \frac{C_U n!}{(p+s) \cdots (p+ns)},$$

where  $C_U$  is the Lebesgue measure of the sub-level set  $\{x \in \mathbb{R}^n : U(x) \leq 1\}$ . We only check the identity for  $s > 0$ , and the other two cases can be proved similarly.

$$\begin{aligned} \int_{\mathbb{R}^n} f_{s,U}(x)^p dx &= p \int_0^1 t^{p-1} |\{x \in \mathbb{R}^n : (1 - sU(x))_+^{1/s} > t\}| dt \\ &= p \int_0^1 t^{p-1} |\{x \in \mathbb{R}^n : U(x) < (1 - t^s)/s\}| dt \\ &= C_U p \int_0^1 t^{p-1} ((1 - t^s)/s)^n dt \\ &= C_U s^{-n-1} p B(p/s, n+1) \end{aligned}$$

In the third equation, we use the homogeneity of  $U$  and the property of Lebesgue measure. Then we can prove the identity using the following fact

$$B(p/s, n+1) = \frac{n!}{(p/s+n) \cdots p/s}.$$

Thus the preceding theorem can be written in the following way: if  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is an integrable  $s$ -concave function, then

$$p \mapsto \frac{\int_{\mathbb{R}^n} f(x)^p dx}{\int_{\mathbb{R}^n} f_{s,U}(x)^p dx}$$

is log-concave for  $p > \max(0, -ns)$ .

## 4 Concentration of information content

Now we are ready to study the concentration property of information content for convex measures introduced and studied by Borell [7, 8].

**Definition 4.1.** Let  $-\infty \leq \kappa \leq \infty$ . A finite Borel measure  $\mu$  on  $\mathbb{R}^n$  is called  $\kappa$ -concave if we have

$$\mu((1-\lambda)A + \lambda B) \geq ((1-\lambda)\mu(A)^\kappa + \lambda\mu(B)^\kappa)^{1/\kappa} \quad (15)$$

for all  $\lambda \in [0, 1]$  and all Borel sets  $A, B \subseteq \mathbb{R}^n$  such that  $\mu(A)\mu(B) > 0$ .

Here  $(1-\lambda)A + \lambda B = \{(1-\lambda)x + \lambda y : x \in A, y \in B\}$  stands for the Minkowski sum of two sets. The limit cases are interpreted by continuity. Thus the right hand side of (15) is equal to  $\min(\mu(A), \mu(B))$  for  $\kappa = -\infty$ ;  $\mu(A)^{1-\lambda}\mu(B)^\lambda$  for  $\kappa = 0$ ; and  $\max(\mu(A), \mu(B))$  for  $\kappa = \infty$ . Note that the inequality (15) becomes stronger as  $\kappa$  increases. For  $\kappa = -\infty$ , we obtain the largest class, whose members are called convex or hyperbolic measures. The case  $\kappa = 0$  describes log-concave measures. For  $\kappa = 1/n$ , up to a factor,  $\mu$  has to be the

Lebesgue measure supported on a convex set of  $\mathbb{R}^n$ . If  $\mu$  is not concentrated at a point, then necessary  $\kappa \leq 1$ . Moreover,  $\mu$  is absolutely continuous with respect to the Lebesgue measure for  $\kappa \leq 1/n$ , and its density function is  $-1/\beta$ -concave with  $\beta = n - 1/\kappa$ .

We say that a  $\mathbb{R}^n$ -valued random variable  $X$  is  $\kappa$ -concave if the probability measure induced by  $X$  is  $\kappa$ -concave. In this section, we let  $X$  be a  $\kappa$ -concave random variable with density  $f$  and  $\kappa < 0$ . Then Borell's characterization implies that there is a convex function  $V$  such that  $f = V^{-\beta}$ . In the following, we will study the deviation of  $\tilde{h}(X)$  from its mean  $h(X)$ , that is corresponding to taking  $\varphi = -\log f$  in Section 2. Then the moment generating function is

$$\mathbb{E}f^{-\alpha}(X) = \int_{\mathbb{R}^n} f(x)^{1-\alpha} dx.$$

The integral is finite as long as  $(1 - \alpha)\beta > n$ , i.e. that  $\alpha < 1 - n/\beta$ .

**Proposition 4.2.** *Let  $\beta > n$  and let  $X$  be a random variable in  $\mathbb{R}^n$  with density  $f$  being  $-1/\beta$ -concave. Then the function*

$$\alpha \mapsto \prod_{i=1}^n ((1 - \alpha)\beta - i) \mathbb{E}f^{-\alpha}(X) \quad (16)$$

*is log-concave for  $\alpha < 1 - n/\beta$ .*

*Proof.* It easily follows from Theorem 3.1 with  $p$  replaced by  $1 - \alpha$  and  $s$  replaced by  $-1/\beta$ .  $\square$

Following Lemma 2.3, we can set

$$c(\alpha) = - \sum_{i=1}^n \log((1 - \alpha)\beta - i). \quad (17)$$

**Corollary 4.3.** *Under the conditions and notations of Proposition 4.2, we have*

$$\text{Var}(\tilde{h}(X)) \leq \beta^2 \sum_{i=1}^n (\beta - i)^{-2}. \quad (18)$$

*Proof.* By Lemma 2.1, we know that  $\text{Var}(\tilde{h}(X)) = L''(\alpha)$ , where  $X_\alpha$  is a random variable with density proportional to  $f^{1-\alpha}$  and  $L(\alpha) = \log \mathbb{E}f^{-\alpha}(X)$  is the logarithmic moment generating function. By Proposition 4.2, we know that  $L''(\alpha) \leq c''(\alpha)$ , where  $c(\alpha)$  is defined in (17). Then the variance bound (18) follows by differentiating  $c(\alpha)$  twice and setting  $\alpha = 0$ .  $\square$

**Remark 4.4.** *The variance bound is sharp. Suppose  $X$  has density  $f = (1 + U/\beta)_+^{-\beta}$  with  $U$  being a positively homogeneous convex function of degree 1. In this case, the function in Proposition 4.2 is log-affine, i.e.  $L''(\alpha) = c''(\alpha)$ . Then we have equality in the above variance bound. In particular, it includes the Pareto distribution with density*

$$f(x) = \frac{1}{Z_n(a, \beta)} (a + x_1 + \cdots + x_n)^{-\beta}, \quad x_i > 0, \quad (19)$$

*where  $a > 0$  and  $Z_n(a, \beta)$  is a normalizing constant.*

Let  $\beta > n + 2$  and let  $X$  be a random variable in  $\mathbb{R}^n$  with density  $f$  being  $-1/\beta$ -concave. In this case, we have  $\mathbb{E}|X|^2 < \infty$  and the covariance matrix  $\Sigma$  is defined by

$$\Sigma = \mathbb{E}(X - \mathbb{E}X) \otimes (X - \mathbb{E}X). \quad (20)$$

Then we have

$$\begin{aligned} n &= \int_{\mathbb{R}^n} \langle x - \mathbb{E}X, -\nabla \log f(x) \rangle f(x) dx \\ &\leq \left( \int_{\mathbb{R}^n} |x - \mathbb{E}X|^2 f(x) dx \cdot \int_{\mathbb{R}^n} |\nabla \log f(x)|^2 f(x) dx \right)^{1/2} \\ &= \sqrt{\text{tr}(\Sigma) J(X)}, \end{aligned}$$

where  $\text{tr}(\Sigma)$  is the trace of  $\Sigma$  and  $J(X)$  is the Fisher information defined by

$$J(X) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} dx. \quad (21)$$

Combining with Corollary 4.3 we have the following result.

**Corollary 4.5.** *Let  $\beta > n + 2$  and let  $X$  be a random variable in  $\mathbb{R}^n$  with density  $f$  being  $-1/\beta$ -concave. Then we have*

$$\text{Var}(\tilde{h}(X)) \leq \frac{\text{tr}(\Sigma)\beta^2}{n^2} \sum_{i=1}^n (\beta - i)^{-2} J(X). \quad (22)$$

In particular, if  $X$  is isotropic, i.e. that  $\mathbb{E}X = 0$  and  $\Sigma$  is the identity matrix, we have

$$\text{Var}(\tilde{h}(X)) \leq \frac{\beta^2}{n} \sum_{i=1}^n (\beta - i)^{-2} J(X). \quad (23)$$

Taking  $\beta \rightarrow \infty$  yields the analogue for log-concave random variables, namely

$$\text{Var}(\tilde{h}(X)) \leq J(X), \quad (24)$$

which was observed by Nguyen [16].

**Theorem 4.1.** *Let  $\beta > n$  and let  $X$  be a random variable in  $\mathbb{R}^n$  with density  $f$  being  $-1/\beta$ -concave. Then we have*

$$\mathbb{E}e^{\alpha(\tilde{h}(X) - h(X))} \leq e^{\psi_c(\alpha)} \quad (25)$$

for  $\alpha < 1 - n/\beta$ , where

$$\psi_c(\alpha) = -\alpha\beta \sum_{i=1}^n (\beta - i)^{-1} - \sum_{i=1}^n \log \frac{(1 - \alpha)\beta - i}{\beta - i}. \quad (26)$$

Particularly, we have equality for Pareto distributions.

*Proof.* The moment generating function bound (25) easily follows from Lemma 2.3 and Proposition 4.2. Some easy calculations will show the equality case for Pareto distributions. Essentially that is due to the identity  $L''(\alpha) = c''(\alpha)$ , where  $c(\alpha)$  is defined in (17).  $\square$



**Corollary 4.6.** *Under the conditions and notations of Theorem 4.1, we have for  $t > 0$  that*

$$\mathbb{P}(\tilde{h}(X) - h(X) > t) \leq e^{-\psi_{c,+}^*(t)}, \quad (27)$$

$$\mathbb{P}(\tilde{h}(X) - h(X) < -t) \leq e^{-\psi_{c,-}^*(-t)}, \quad (28)$$

where  $\psi_{c,+}^*$  and  $\psi_{c,-}^*$  are Fenchel-Legendre dual functions of  $\psi_{c,+}$  and  $\psi_{c,-}$ , respectively.

In general we do not have explicit expressions for  $\psi_{c,+}^*$  or  $\psi_{c,-}^*$ . The following result was obtained by Bobkov and Madiman [3] with the assumption  $\beta \geq n + 1$ , which can be relaxed to  $\beta > n$ . It basically says that the entropy of an  $\kappa$ -concave distribution can not exceed that of the Pareto distribution with the same maximal density value.

**Corollary 4.7.** *Under the conditions and notations of Theorem 4.1, we have*

$$h(X) \leq -\log \|f\|_\infty + \beta \sum_{i=1}^n (\beta - i)^{-1}, \quad (29)$$

where we denote by  $\|f\|_\infty$  the essential supremum. We have equality for Pareto distributions.

*Proof.* As a function of  $\alpha$ , we have

$$(-\alpha t - \psi_c(\alpha))' = -t + \beta \sum_{i=1}^n (\beta - i)^{-1} - \beta \sum_{i=1}^n ((1 - \alpha)\beta - i)^{-1}.$$

For any  $t > \beta \sum_{i=1}^n (\beta - i)^{-1}$ , we can see that  $-\alpha t - \psi_c(\alpha)$  is a decreasing function of  $\alpha < 1 - n/\beta$ . It is clear that  $\lim_{\alpha \rightarrow -\infty} (-\alpha t - \psi_c(\alpha)) = \infty$ . Therefore we have  $\psi_{c,-}^*(-t) = \infty$  for  $t > \beta \sum_{i=1}^n (\beta - i)^{-1}$ . Using the lower tail estimate in Corollary 4.6, almost surely we have

$$\tilde{h}(X) - h(X) \geq -\beta \sum_{i=1}^n (\beta - i)^{-1}.$$

Taking the supremum over all realizable values of  $X$  yields

$$-\log \|f\|_\infty - h(X) \geq -\beta \sum_{i=1}^n (\beta - i)^{-1}.$$

That is equivalent to the desired statement.  $\square$

**Remark 4.8.** *We can get corresponding estimates for log-concave random variables (see [3, 11]) by taking the limit  $\beta \rightarrow \infty$ .*

The following result is an improvement of Proposition 5.1 of Bobkov and Madiman [4]. Its analogue for log-concave probability measures was first observed by Klartag and Milman [14], with refinement made by [11, Corollary 4.7].

**Corollary 4.9.** *Under the conditions and notations of Theorem 4.1, we have*

$$\mathbb{P}(f(X) \geq c_0^n \|f\|_\infty) \geq 1 - c_1^n \quad (30)$$

for  $0 < c_0 < 1$  such that  $n \log c_0 < -\beta \sum_{i=1}^n (\beta - i)^{-1}$  and some  $0 < c_1 < 1$  depending on  $c_0$  and  $\beta$ .

*Proof.* Note that

$$\begin{aligned}
\mathbb{P}(f(X) \leq c_0^n \|f\|_\infty) &= \mathbb{P}(\log f(X) \leq \log \|f\|_\infty + n \log c_0) \\
&= \mathbb{P}(\tilde{h}(X) \geq -\log \|f\|_\infty - n \log c_0) \\
&\leq \mathbb{P}\left(\tilde{h}(X) \geq h(X) - \beta \sum_{i=1}^n (\beta - i)^{-1} - n \log c_0\right)
\end{aligned}$$

We use Corollary 4.7 in the above inequality. Applying the upper tail estimate of Corollary 4.6 with

$$t = -n \log c_0 - \beta \sum_{i=1}^n (\beta - i)^{-1} \quad (31)$$

yields

$$\mathbb{P}(f(X) \leq c_0^n \|f\|_\infty) \leq e^{-\psi_{c,+}^*(t)}. \quad (32)$$

As a function of  $\alpha$ , we have

$$(\alpha t - \psi_c(\alpha))' = -n \log c_0 - \sum_{i=1}^n \frac{\beta}{(1 - \alpha)\beta - i},$$

from which we can see  $(\alpha t - \psi_c(\alpha))'(0) = t > 0$  and  $(\alpha t - \psi_c(\alpha))'(1 - n/\beta) = -\infty$ . In addition we can see the concavity of  $\alpha t - \psi_c(\alpha)$  from

$$(\alpha t - \psi_c(\alpha))'' = - \sum_{i=1}^n \frac{\beta^2}{((1 - \alpha)\beta - i)^2} < 0.$$

Therefore we have

$$\psi_{c,+}^*(t) = \alpha^* t - \psi_c(\alpha^*), \quad (33)$$

where  $\alpha^*$  is a positive number such that  $(\alpha t - \psi_c(\alpha))'(\alpha^*) = 0$ , i.e. that

$$\sum_{i=1}^n \frac{\beta}{(1 - \alpha^*)\beta - i} = -n \log c_0. \quad (34)$$

Using the definitions of  $\psi_c(\alpha)$  and  $t$  in (26) and (31), respectively, we have

$$\psi_{c,+}^*(t) = -n\alpha^* \log c_0 + \sum_{i=1}^n \log \frac{(1 - \alpha^*)\beta - i}{\beta - i}. \quad (35)$$

Combining with (32), we have

$$\mathbb{P}(f(X) \leq c_0^n \|f\|_\infty) \leq c_1^n, \quad (36)$$

where

$$c_1 = c_0^{\alpha^*} \left( \prod_{i=1}^n \frac{\beta - i}{(1 - \alpha^*)\beta - i} \right)^{1/n}. \quad (37)$$

That is equivalent to the desired statement. To see that  $c_1 < 1$ , we take the logarithm of  $c_1$ ,

$$\begin{aligned}
\log c_1 &= \alpha^* \log c_0 + \frac{1}{n} \sum_{i=1}^n \log \frac{\beta - i}{(1 - \alpha^*)\beta - i} \\
&= -\frac{1}{n} \sum_{i=1}^n \frac{\alpha^* \beta}{(1 - \alpha^*)\beta - i} + \frac{1}{n} \sum_{i=1}^n \log \frac{\beta - i}{(1 - \alpha^*)\beta - i} \\
&= -\frac{1}{n} \sum_{i=1}^n \left( \frac{\alpha^* \beta}{(1 - \alpha^*)\beta - i} - \log \left( 1 + \frac{\alpha^* \beta}{(1 - \alpha^*)\beta - i} \right) \right) \\
&< 0.
\end{aligned}$$

We use the equation (34) in the second identity. The last inequality follows from the fact that  $\log(1 + x) < x$  for  $x > 0$ .  $\square$

**Acknowledgments:** M.F. was supported in part by the Agence Nationale de la Recherche, project GeMeCoD (ANR 2011 BS01 007 01), J.L. and M.M. were supported in part by the U.S. National Science Foundation through grants DMS-1409504 and CCF-1346564.

## References

- [1] S. Bobkov, M. Fradelizi, J. Li, and M. Madiman. When can one invert Hölder’s inequality? (and why one may want to). *Preprint*, 2015.
- [2] S. Bobkov and M. Madiman. Concentration of the information in data with log-concave distributions. *Ann. Probab.*, 39(4):1528–1543, 2011.
- [3] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, 2011.
- [4] S. Bobkov and M. Madiman. Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.*, 262(7):3309–3339, 2012.
- [5] F. Bolley, I. Gentil, and A. Guillin. Dimensional improvements of the logarithmic sobolev, talagrand and brascamp-lieb inequalities. *Preprint*, 2015.
- [6] C. Borell. Complements of Lyapunov’s inequality. *Math. Ann.*, 205:323–331, 1973.
- [7] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12:239–252, 1974.
- [8] C. Borell. Convex set functions in  $d$ -space. *Period. Math. Hungar.*, 6(2):111–136, 1975.
- [9] T. M. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inform. Theory*, 35(1):37–43, 1989.
- [10] M. Fradelizi, O. Guédon, and A. Pajor. Thin-shell concentration for convex measures. *Studia Math.*, 223(2):123–148, 2014.
- [11] M. Fradelizi, M. Madiman, and L. Wang. Optimal concentration of information content for log-concave densities. *Preprint*, 2015.

- [12] O. Guédon, P. Nayar, and T. Tkocz. Concentration inequalities and geometry of convex bodies. In *Analytical and probabilistic methods in the geometry of convex bodies*, volume 2 of *IMPAN Lect. Notes*, pages 9–86. Polish Acad. Sci. Inst. Math., Warsaw, 2014.
- [13] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.
- [14] B. Klartag and V. D. Milman. Geometry of log-concave functions and measures. *Geom. Dedicata*, 112:169–182, 2005.
- [15] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms*, 4(4):359–412, 1993.
- [16] V. H. Nguyen. Dimensional variance inequalities of brascamp-lieb type and a local approach to dimensional prékopa’s theorem. *J. Funct. Anal.*, 266:931–955, 2014.
- [17] V. H. Nguyen. Inégalité fonctionnelles et convexité. *PhD thesis, Université Pierre et Marie Curie (Paris VI)*, October 2013.
- [18] L. Wang. Heat capacity bound, energy fluctuations and convexity. *PhD thesis, Yale University*, May 2014.